

Equitability of Dependence Measure

Hangjin Jiang^{1,2} Kan Liu¹ Yiming Ding^{1*}

1. Wuhan institute of Physics and Mathematics, Chinese Academy of Sciences, Wuhan, China

2. University of Chinese Academy of Sciences, Beijing, China

Abstract A measure of dependence is said to be equitable if it gives similar scores to equally noisy relationship of different types. In practice, we do not know what kind of functional relationship is underlying two given observations, Hence the equitability of dependence measure is critical in analysis and by scoring relationships according to an equitable measure one hopes to find important patterns of any type of further examination. In this paper, we introduce our definition of equitability of a dependence measure, which is naturally from this initial description, and Further more power-equitable(weak-equitable) is introduced which is of the most practical meaning in evaluating the equitability of a dependence measure.

Keywords: Dependence, mutual information, Equitability, weak-equitability.

1 Introduction

Measuring dependence between two variables or sequences plays a fundamental role in various kind of data analysis, such as fMRI data, genetic data. How should one quantify such a dependence without bias for relationship of a specific form? This gives rise to the concept "Equitability" of a dependence measure [10], [36]. By scoring relationships according to an equitable measure one hopes to find important patterns of any type further examination. The first description of "Equitability" given in [10] is "A measure of dependence is said to be equitable if it gives similar scores to equally noisy relationship of different types". In [37], authors pointed out that there is no dependence satisfying the definition of equitability given in [10], the R^2 -Equitability. Furthermore, they give a definition of "self-equitability", and show by theoretically proof, and numerical simulation that mutual information satisfies the "self-equitability".

Equitability of dependence measure means it is equitable to all kinds of functional relationship, as the description goes, give similar score to equally noisy functional relationship. Taking correlation for example, it is un-equitable, since it gives very small scores to nonlinear relationships that can not be well approximated by linear function. There are two key terms in this sentence "equally noisy relationship" and "similar score". The noise level is given by $1 - R^2\{f(x), y\}$ for the model $y = f(x) + \epsilon$ in [36], and [37]. However, in their definition of R^2 -Equitability there is no words related to "equally noisy relationship". In this paper, we introduce the model signal-to-noise ratio (MSNR) and signal signal-to-noise ratio (SSNR) to control the noise level in our simulation, and give the definition of Equitability in a more natural way, which, at the same time, leads to the definition of weak-equitable.

The self-equitable defined and discussed in [37] focused on the equitability of dependence measure when it is regarded as a measure of noise in the data set. If a dependence measure is self-equitable, then it does not depend on what the specific functional relationship between x and y would be in the absence of noise. However, the noise in the data set is difficult to measure, or no one can tell though the methods existed how much noise is in the data set. In other words, self-equitable is necessary for a dependence measure used as a noise measure. However, in practice, people usually used the dependence measure as a test statistic, we say x and y are significantly associated with each other when the value $D[x; y]$ obtained by dependence measure D is larger than a given threshold, λ , with significant value α , which motivated us to the definition of power-equitable, that is, weak-equitable.

In this paper, we try to make clear the statement debates in [39] that MIC is more equitable than MI. In addition, we introduce a new definition, weak-equitable (power-equitable), which is meaningful when a dependence measure is used in independent test problem. In the next section, we will recall firstly the

*Correspondence author: ding@wipm.ac.cn

definitions given in [37], and then introduce the definition of equitable, and weak-equitable. Simulation results are given in section 3.

2 Definitions of Equitability

A measure of dependence is said to be equitable if it gives similar scores to equally noisy relationships of different types [10],[36]. In other words, a measure of how much noise is in an x-y scatter plot should not depend on what the specific functional relationship between x and y would be in the absence of noise [37]. Justin B. Kinney et al. gave the definition of R^2 -equitable and Self-equitable [37]. They pointed out that the dependence measure satisfying R^2 -equitable does not exist, and dependence measure satisfying Self-equitability exists, mutual information is one of them.

Here we recall the definition of Self-equitable and R^2 -equitable.

Definition 2.1 A dependence measure $D[X;Y]$ is R^2 -equitable if and only if, when evaluated on a joint probability distribution $P(X,Y)$, that corresponds to a noisy functional relationship between two real random variables X and Y , the following relation holds:

$$D[X;Y] = g(R^2[f(X);Y]) \quad (1)$$

Here, g is a function that does not depend on $P(X,Y)$ and f is the function defining the noisy functional relationship, i.e., $Y = f(X) + \eta$, for some random variable η . The noise term η may depend on x as long as $x \rightarrow f(x) \rightarrow \eta$ forms a Markov chain.

Definition 2.2 A dependence measure $D[x;y]$ is self-equitable if and only if

$$D[x;y] = D[f(x);y] \quad (2)$$

whenever f is a deterministic function and $x \rightarrow f(x) \rightarrow y$ forms a Markov chain.

In the definition of R^2 -equitable, the term "gives similar scores to equally noisy relationships of different types" is described by the dependence measure as a function, independent with x and y , of noise, so the meaning of equitability is conveyed implicitly. Differently, our definition of equitable conveys the meaning "equally noisy relationship", directly through a mathematical definition of noisy-equal model based on the model signal-to-noise ratio(MSNR).

Definition 2.3 (MSNR) Signal-to-Noise Ration(SNR) for a model, $y = f(x) + \varepsilon$, is given by

$$MSNR_\varepsilon(f) = \frac{\text{var}(y)}{\text{var}(\varepsilon)} \quad (3)$$

where $\text{var}(x)$ is the variance of x .

Remark 1:

1. Two models, $y_1 = f_1(x) + \varepsilon_1$ and $y_2 = f_2(x) + \varepsilon_2$ with the same MSNR is called **noisy-equal models**.
2. In [10], [37], the noise level is measured by $1 - R^2(f(x), f(x) + \varepsilon) = 1 - \rho^2(f(x), f(x) + \varepsilon)$, theoretically, we have

$$R^2(f(x), f(x) + \varepsilon) = \frac{1}{\sqrt{1 + \frac{1}{MSNR_\varepsilon(f)}}} \quad (4)$$

3. How to get noisy-equal models? Given two models, $y_1 = f_1(x) + \varepsilon_1$ and $y_2 = f_2(x) + \varepsilon_2$, we set $\varepsilon_1 = \varepsilon$, $\varepsilon_2 = \frac{\varepsilon}{a}$, where $a^2 = \frac{\text{var}(f_1(x))}{\text{var}(f_2(x))}$, ε is independent with x , then

$$\begin{aligned}
MSNR_{\varepsilon_1}(f_1) &= \frac{var(y_1)}{var(\varepsilon)} = \frac{var(f_1(x)) + var(\varepsilon)}{var(\varepsilon)} \\
&= \frac{var(f_1(x))/a^2 + var(\varepsilon/a)}{var(\varepsilon/a)} \\
&= \frac{var(y_2)}{var(\varepsilon_2)} = MSNR_{\varepsilon_2}(f_2)
\end{aligned} \tag{5}$$

Hence, we get noisy-equal models $y_1 = f_1(x) + \varepsilon$ and $y_2 = f_2(x) + \frac{\varepsilon}{a}$, where $a^2 = \frac{var(f_1(x))}{var(f_2(x))}$.

However, when mean value is not zero, $MSNR_{\varepsilon}$ is not equal to SNR, we need to go back to the definition of real SNR, which is $SNR_{\varepsilon}(f) = \frac{Power(y)}{Power(\varepsilon)}$. In most of the case, especially in electronic engineers, it equal to the signal square to noise square ration.

Definition 2.4 (SSNR) SNR for signals, $y = f(x) + \varepsilon$, is given by

$$SSNR_{\varepsilon}(f) = \frac{y^2}{\varepsilon^2} \tag{6}$$

Remark 2:

1. Two signals, $y_1 = f_1(x) + \varepsilon_1$ and $y_2 = f_2(x) + \varepsilon_2$ with the same SSNR is called **noisy-equal signals**.
2. We use a heuristic algorithm to get noisy-equal signals. Given two signals, $y_1 = f_1(x) + \varepsilon_1$ and $y_2 = f_2(x) + \varepsilon_2$, we calculate the SSNR of y_1 , use it as optimization objective. Then we generate random ε_2 , and let SSNR of y_2 approximate SSNR of y_1 .

Algorithm 1: The heuristic algorithm for generating noisy-equal signals

```

1  snr.o.r <- function(sig, noise){
2      return (crossprod(sig, sig)/
              crossprod(noise, noise))
3  }
4
5  guessNoise.r <- function(sig, snr,
    steps=100, thr=0.03){
6      crossprod_noise <- crossprod(sig,
    sig)/snr
7      n <- length(sig)
8      rgnoise <- rep(0, n)
9      step <- 0
10     delta <- 9999
11     an <- (crossprod_noise/n)^0.5
12     mxan <- an
13     mian <- mxan*-1
14     snr.r <- 0
15     repeat{
16         mean_noise <- runif(1, min=mian,
            max=mxan)
17         sd_noise <- (crossprod_noise/n-

```

Then, we fix $n-1$ components of ε_2 , we get

$$SSNR_{\varepsilon}(f_1) = \frac{\sum_{i=1}^{n-1} (f_2(x_i) + \varepsilon_{2i})^2 + (f_2(x_n) + \varepsilon_{2n})^2}{\sum_{i=1}^{n-1} \varepsilon_{2i}^2 + \varepsilon_{2n}^2} \tag{7}$$

Because $SSNR_\varepsilon(f_1)$, $f_2(x_i)$, $i = 1, \dots, n$ and ε_{2i} , $i = 1, \dots, n - 1$ are all known, we can solve the eq.7, and generate accurate noise to satisfy the SSNR.

Definition 2.5 (*Equitable*) A dependence measure $D[x; y]$ is *Equitable* if and only if

$$D[x; y_1] = D[x; y_2] \quad (8)$$

for any noisy-equal models $y_1 = f_1(x) + \varepsilon_1$ and $y_2 = f_2(x) + \varepsilon_2$.

This definition is theoretically equal to that of R^2 -equitable presented in [37], but from a different point of view, which is more naturally and heuristically. Further, it helps us to look great insight into equitability of dependence measures, seen in the following section.

The self-equitable focused on the equitability of dependence measure when it is regarded as a measure of noise in the data set. If a dependence measure is self-equitable, then it does not depend on what the specific functional relationship between x and y would be in the absence of noise. However, the noise in the data set is difficult to measure, or no one can tell though the methods existed how much noise is in the data set. In other words, self-equitable is necessary for a dependence measure used as a noise measure. However, in practice, people usually used the dependence measure as a test statistic, we say x and y are significantly associated with each other when the value $D[x; y]$ obtained by dependence measure D is larger than a given threshold, λ , with significant value α , which motivated us to the definition of power-equitable, that is, weak-equitable. To be formally,

Definition 2.6 (*Power-equitable*) A dependence measure $D[x; y]$ is *Power-equitable* if and only if

$$Power[f_1] = Power[f_2] \quad (9)$$

for any noisy-equal models $y_1 = f_1(x) + \varepsilon_1$ and $y_2 = f_2(x) + \varepsilon_2$, where $Power[f]$ is the testing power of dependence measure D for detecting functional relationship f underlying noisy data set.

So, if a dependence measure is equitable, then it is power-equitable, but the converse is not right, and we refer the power-equitable as weak-equitable.

Weak-equitability can be explained as that: if a dependence measure, D , is weak-equitable, then all kinds of functional relationship underlying equally noisy data sets will be detected with the same possibility by D with significant level α , which is quite meaningful and interesting. This allowed, similar to a equitable case, us to use D as a test to determine whether or not x and y is significantly associated with each other. As mentioned before, self-equitable is meaningful, when D is used as a noise measure.

3 Simulation Results

In this section, we discuss the self-equitability, equitability, and power-equitability of the most popular dependence measures: MIC[10], CDC [38], RDC[15], HHG[5], Pearson Correlation coefficient(pcor), Spearman's Rank Correlation(scor), Kentall's τ (kcor), curve correlation[13], HSIC [27], normalized mutual information(MI) [10]. A detailed discussion of these measures can found in [38]. Although the Equitablity of dcor [4] is discussed in [36], we also discuss it. In [36], authors pointed that there is a trade-off between equitability and test power of MIC. We will also discuss it in the following section.

3.1 Equitability

In this part, we analysis the equitability of MIC, CDC, RDC, HHG, pcor, scor, kcor, dcor, HSIC and MI. Although, there is some theoretical results given in [37] showing that dependence measure satisfying the definition of equitable do not exist, we give a detailed simulation results to further explanation.

In our simulation, each time x is sampled from uniform distribution with length 5000, and noise is drawn from standard normal distribution, the MSNR is controlled by using (5), the SSNR is controlled by the heuristic algorithm, then we get one value of $D(x, y = f(x) + \varepsilon)$. We do this 100 times for each kind of functional types defined in appendix.

In Figure 1, Equitability of ACE, and RDC is analyzed. According to the definition of equitability (Definition 2.6), if we get almost the same score given by dependence measure D to each kind of functional type, then D is equitable. In our simulation, it is equal to that if the lines with different colors in the figure are consistent, then D is equitable. From the results presented, we can see that both RDC and ACE are not equitable.

In Figure 2, Equitability of MIC, MI, CDC, poor, kcor, scor is analyzed, and all of them are not equitable. Interestingly, MI is equitable in a small set of functional types except four types: Sigmoid, Lopsided L-shaped, L-shaped and Spike. MI is more equitable than MIC, since MI values of 17 functional types (except the four types mentioned before) are centered around 0.8 with smaller variance than that of MIC.

We can find that dependence measures in Figure 3 are not as stable as that of Figure 1 and Figure 2. However the $\text{sd}(\text{SSNR})$ and $\text{range}(\text{SSNR})$ are not larger than $\text{sd}(\text{MSNR})$ and $\text{range}(\text{MSNR})$ correspondingly. It because the heuristic algorithm 1 and eq. 7 used by SSNR simulations random generate the noise ε_2 , but eq. 5 generate the noise ε_2 based on ε_1 .

3.2 Weak-Equitable

In this part, we discuss the weak-equitable of these dependence measures. In our simulation, we set MSNR ranging from 0.7509836 to 3.0057, because high power in models with large MSNR is basic requirement for dependence measure using in independent test, and it is inevitable that for all dependence measure it will have a very lower power on heavy noisy models (models with very small MSNR). The results are given in Figure 4, Figure 5, and Figure 6.

From the results presented in the following figures, we know that HHG is almost power-equitable, but to get its value is time-consuming in large data sets, ACE is secondary to HHG. But ACE is sensitive to outliers, so, we recommend to use CDC in large data sets as dependence measure.

4 Conclusion

In this paper, we discussed the equitability and weak-equitability (power-equitability) of some common dependence measures, such as MIC, MI, dcor, CDC, ACE, RDC, HHG, Pearson correlation coefficient (pcor), Spearman rank correlation (scor), kendall's τ , HSIC. The equitability requires a dependence measure gives similar scores to equally noisy data sets no matter what kind of functional relationship is underlying them, this requirement is so strong that it is can not be satisfied by any dependence measure. Therefore, we give our definition of power-equitability, which only requires that all kinds of functional relationship, linear or non-linear, can be detected with the same possibility by dependence measure D at significant level α , when it is used in a independent test with the null hypothesis: x and y are independent, it is called also as weak-equitability because if D is equitable, then it is power-equitable.

Self-equitable is meaningful and the basic requirement for a dependence measure used as a noise measure. However, a self-equitable measure, such as MI, may not have a higher testing power as shown in our simulation results (see Figure 5).

Based on our simulation, we find that MI is more equitable than MIC, and HHG is power-equitable, ACE and CDC is secondary to HHG. Despite the computation time, HHG is worthing highly recommendation in each kind of application.

References

- [1] A.Renyi. On Measures of Dependence. Acta Math.Acad.Sci. Hungar, 10:441-451, 1959.
- [2] A.Kolmogorove. Grundbegriffe der wahrcheinlichkeitsrechnung, Ergebnisse der(1933) Math.u. Grenzgebiete, Berlin
- [3] L.M.Gelfand, A.M. Jaglom and A.N.Kolmogorove,Zur allgemeinen Definition der Information. Arbeiten zur Informationstheorie, 1959, p57-60.
- [4] Gabor Szekely and Maria L.rizzo. Brownian Distance Covariance. The Annals of applied statistics 2009. Vol.3 No.4 1236-1265

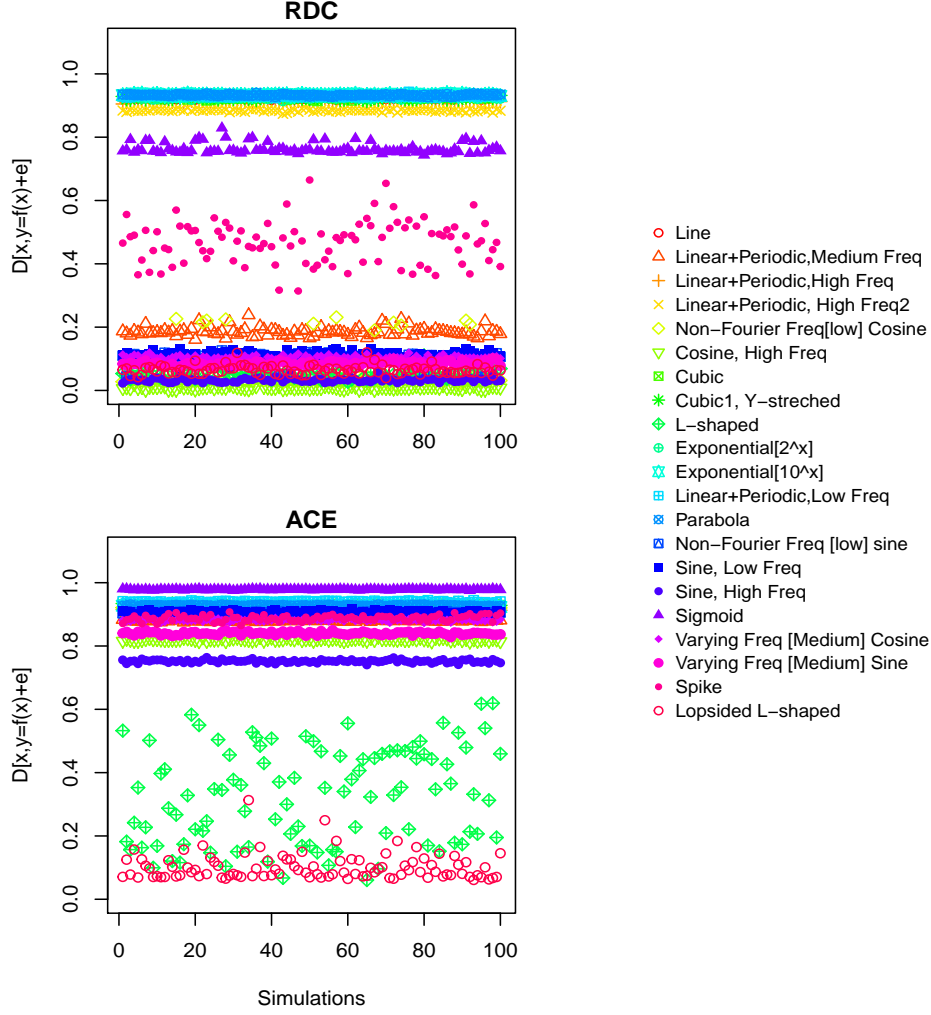


Figure 1: (color online) Equitability of ACE and RDC with $\text{mean}(\text{MSNR})=11.529$, $\text{sd}(\text{MSNR})=0.233$, $\text{range}(\text{MSNR})$ is 10.678 to 12.471. The simulation is based on 21 types of relationship (Definitions are given in appendix), and MSNR is controlled using (5). In this simulation we take the sample x from uniform distribution with length 5000, and noise is sampled from standard normal distribution. We sample x 100 times and get 100 values of dependence measure for each kind of functional type. According to the definition of equitability, if we get almost the same score given by dependence measure D to each kind of functional type, then D is equitable. In our simulation, it is equal to that if the lines with different colors in the figure are consistent, then D is equitable. we can see from the figure they are both not equitable.

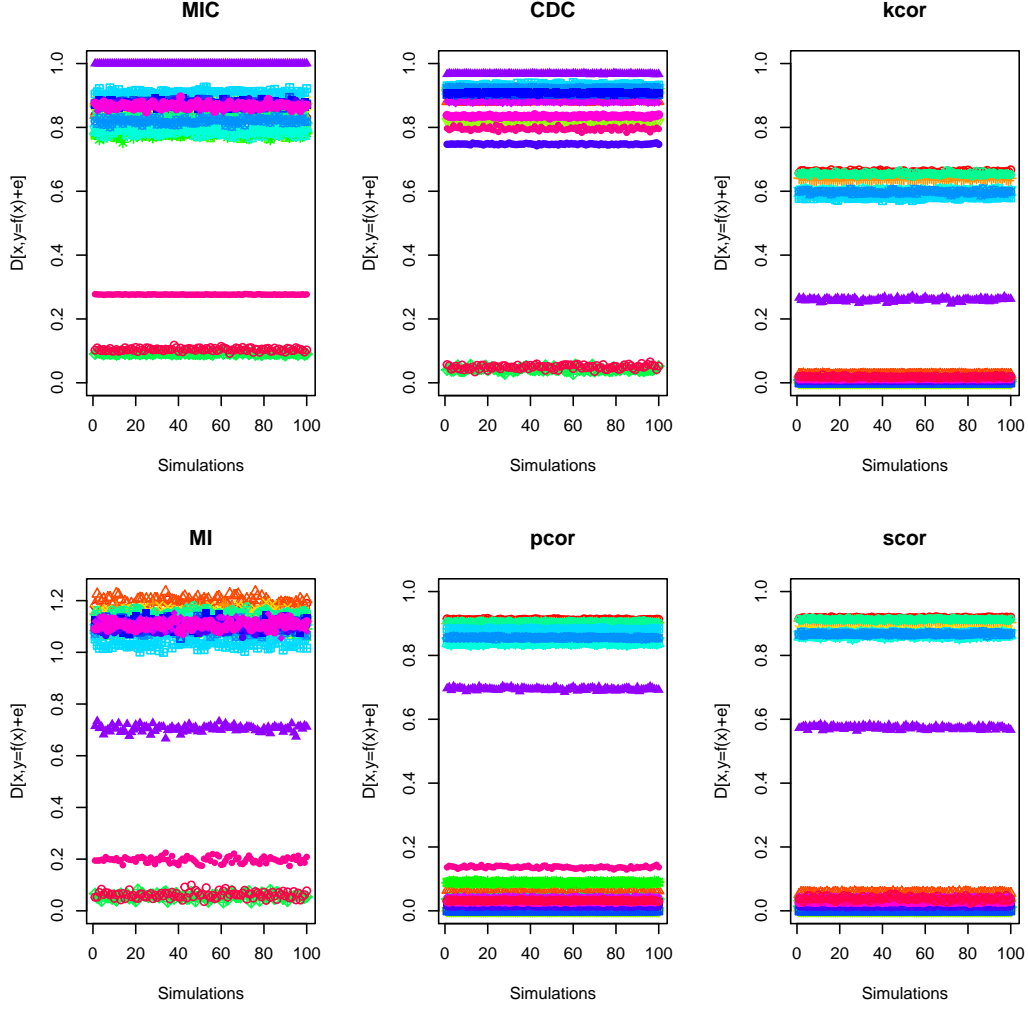


Figure 2: (color online) Equitability of MIC, MI, CDC, pcor, kcor, scor with $\text{mean}(\text{MSNR})=11.529$, $\text{sd}(\text{MSNR})=0.233$, $\text{range}(\text{MSNR})$ is 10.678 to 12.471. (Legends are the same as that in Figure 1.) The simulation is based on 21 types of relationship (Definitions are given in appendix), and MSNR is controlled using (5). In this simulation we take the sample x from uniform distribution with length 5000, and noise is sampled from standard normal distribution. We sample x 100 times and get 100 values of dependence measure for each kind of functional type. According to the definition of equitability, under these setting, if we get almost the same score given by dependence measure D to each kind of functional type, then D is equitable. we can see from the figure they are all not equitable. In our simulation, it is equal to that if the lines with different colors in the figure are consistent, then D is equitable. MI is equitable in a small set of functional types except four types: Sigmoid, Lopsided L-shaped, L-shaped and Spike. MI is more equitable than MIC, since MI values of 17 functional types (except the four types mentioned before) are centered around 1.0 to 1.2 with smaller variance than that of MIC.

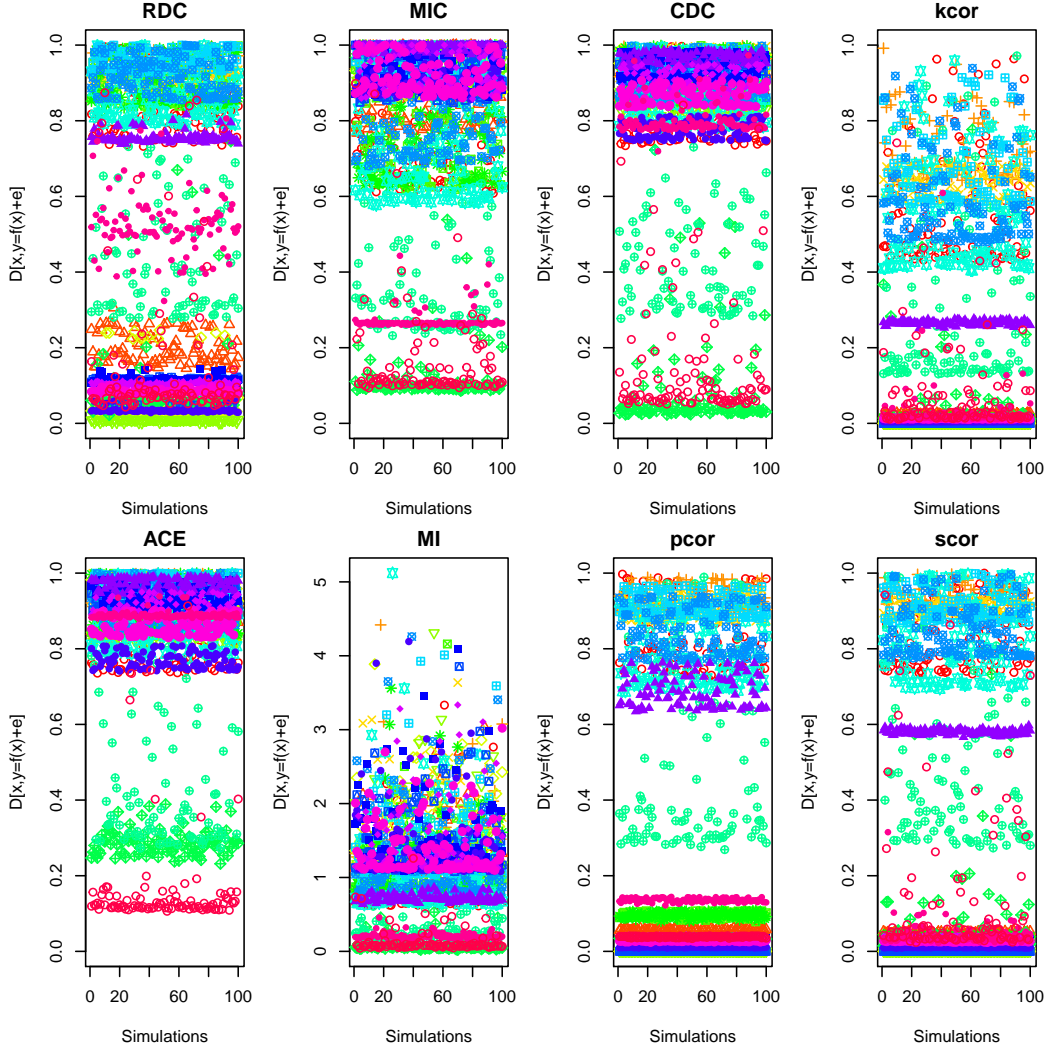


Figure 3: (color online) Equitability of RDC, ACE, MIC, MI, CDC, pcor, kcor, scor with $\text{mean}(\text{SSNR})=10.471$, $\text{sd}(\text{SSNR})=0.209$, $\text{range}(\text{SSNR})$ is 9.382 11.081. (Legends are the same as that in Figure 1.) The simulation is based on 21 types of relationship (Definitions are given in appendix), and SSNR is controlled by the heuristic algorithm 1 and eq. 7. In this simulation we take the sample x from uniform distribution with length 5000, and noise is sampled from standard normal distribution. We sample x 100 times and get 100 values of dependence measure for each kind of functional type. According to the definition of equitability, under these setting, if we get almost the same score given by dependence measure D to each kind of functional type, then D is equitable. we can see from the figure they are all not equitable. In our simulation, it is equal to that if the lines with different colors in the figure are consistent, then D is equitable. MIC is equitable in a large set of functional types except four types: Exponential[2^x], Lopsided L-shaped, L-shaped and Spike. MI is less equitable than MIC, since MIC values of 17 functional types (except the four types mentioned before) are centered around 0.8 with smaller variance than that of MI.

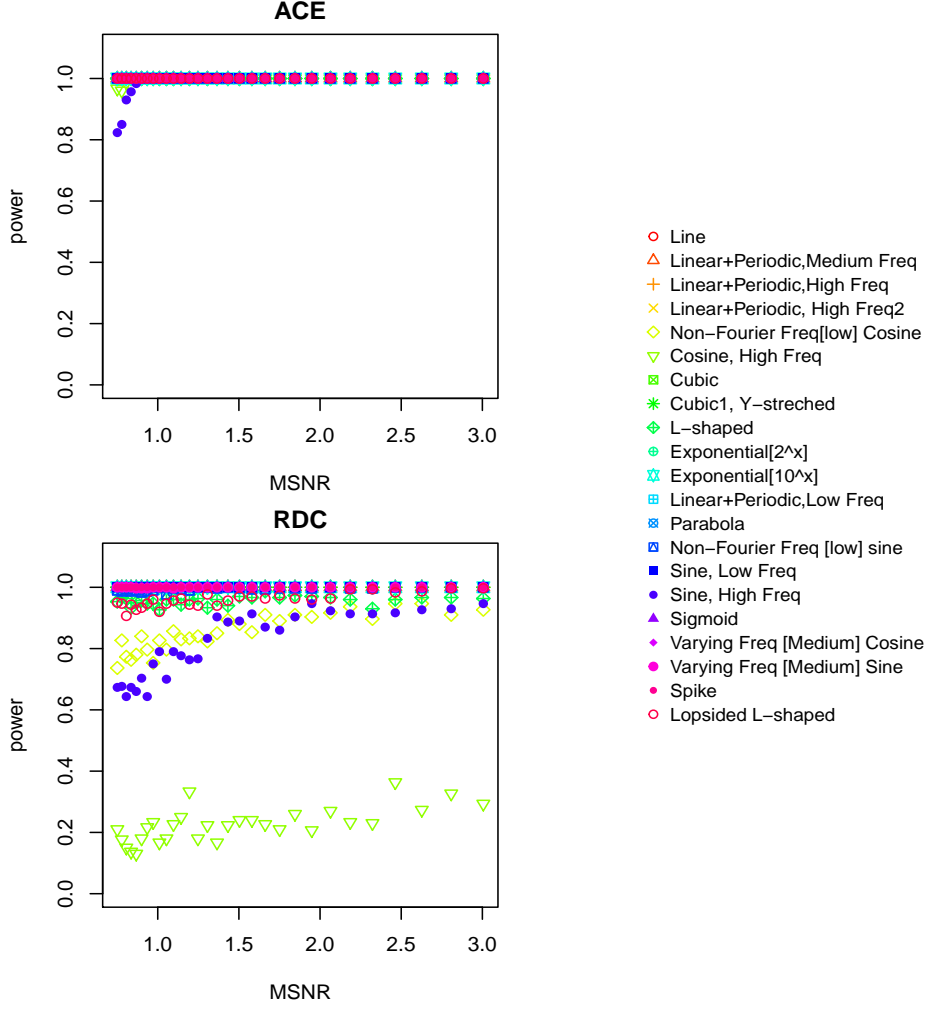


Figure 4: (color online) Weak-Equitale of ACE and RDC with mean MSNR range from 0.7509836 to 3.0057, and standard variance of MSNRs range from 0.039 to 0.15947. (Legends are the same as that in Figure 1.) The simulation is based on 21 types of relationship (Definitions are given in appendix), and MSNR is controlled using (5). In this simulation we take the sample x from uniform distribution with length 5000, and noise is sampled from standard normal distribution. The power was estimated from 300 times simulation at significant level $\alpha = 0.05$. We can see from the results that ACE is much more equitable than RDC in that range of MSNR, here we emphasize on the range of MSNR because if there is a very small MSNR for the data, then all of the methods will have a very small power.

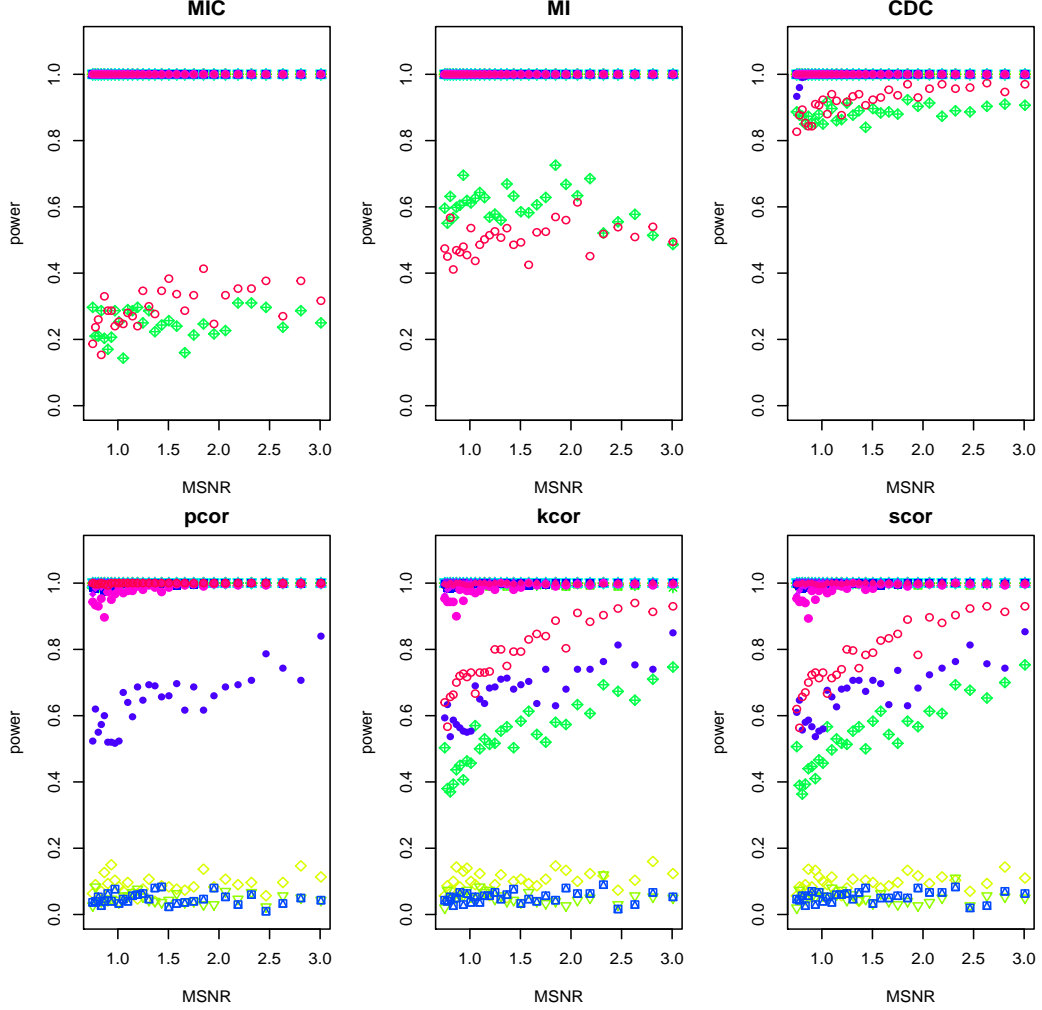


Figure 5: (color online) Weak-Equitable of MIC, MI, CDC, pcor, kcor and scor with mean MSNR range from 0.7509836 to 3.0057, and standard variance of MSNRs range from 0.039 to 0.15947. (Legends are the same as that in Figure 1.) The simulation is based on 21 types of relationship (Definitions are given in appendix), and MSNR is controlled using (5). In this simulation we take the sample x from uniform distribution with length 5000, and noise is sampled from standard normal distribution. The power was estimated from 300 times simulation at significant level $\alpha = 0.05$. We can see from the results that CDC is much more equitable than others in that range of MSNR, here we emphasize on the range of MSNR because if there is a very small MSNR for the data, then all of the methods will have a very small power. For MI and MIC, their power-equitable is very similar, however, it seems that MI is more powerful than MIC in these functional types.

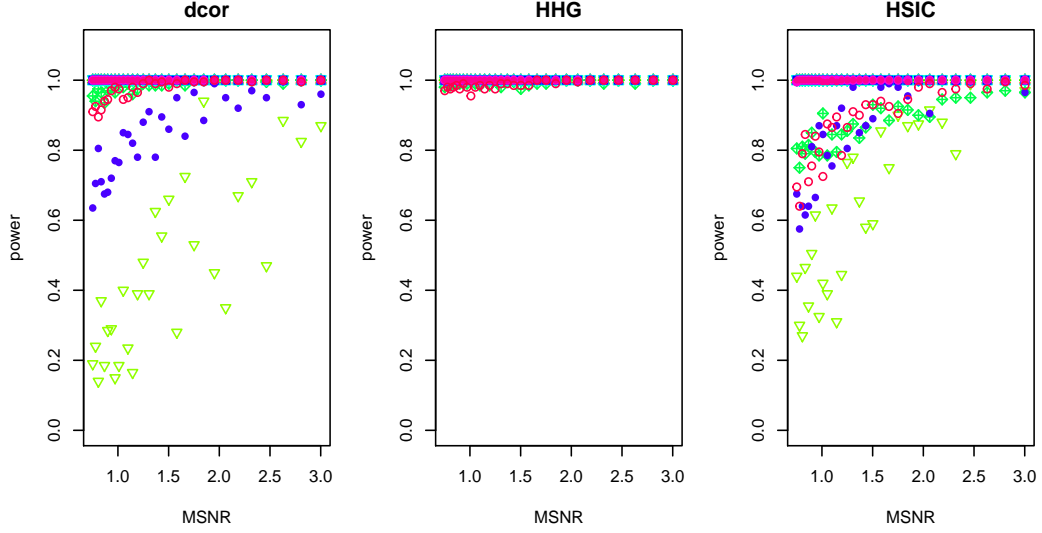


Figure 6: (color online) Weak-Equitale of dcor, HHG, and HSIC with mean MSNR range from 0.7509836 to 3.0057, and standard variance of MSNRs range from 0.039 to 0.15947. (Legends are the same as that in Figure 1.) The simulation is based on 21 types of relationship (Definitions are given in appendix), and MSNR is controlled using (5). In this simulation we take the sample x from uniform distribution with length 5000, and noise is sampled from standard normal distribution. The power was estimated from 300 times simulation at significant level $\alpha = 0.05$. We can see from the results that HHG is much more equitable than others in that range of MSNR, here we emphasize on the range of MSNR because if there is a very small MSNR for the data, then all of the methods will have a very small power.

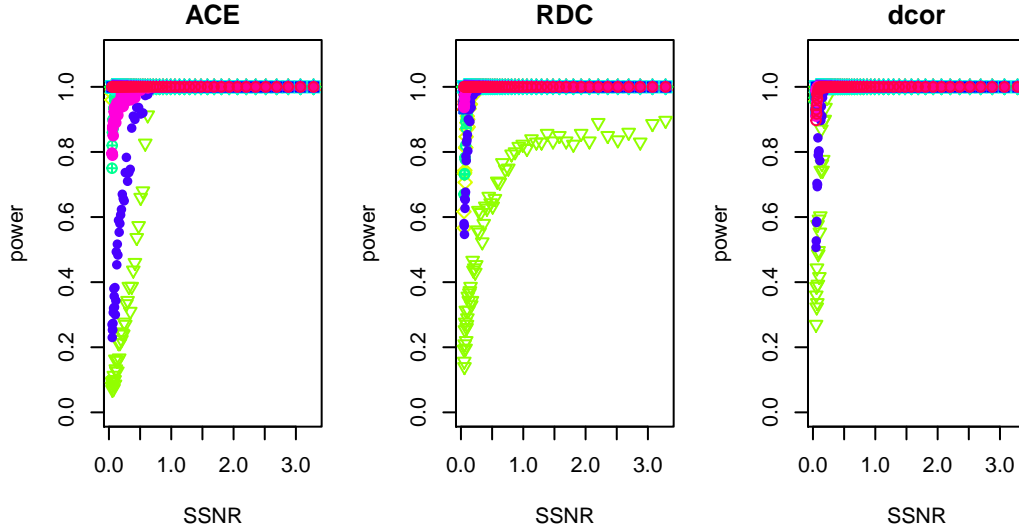


Figure 7: (color online) Weak-Equitale of ACE, RDC and dcor with mean SSNR range from 0.0496 to 3.282. (Legends are the same as that in Figure 1.) The simulation is based on 21 types of relationship (Definitions are given in appendix), and SSNR is controlled by the heuristic algorithm 1 and eq. 7. In this simulation we take the sample x from uniform distribution with length 5000, and noise is sampled from standard normal distribution. The power was estimated from 300 times simulation at significant level $\alpha = 0.05$. We can see from the results that dcor is much more equitable than others in that range of SSNR.

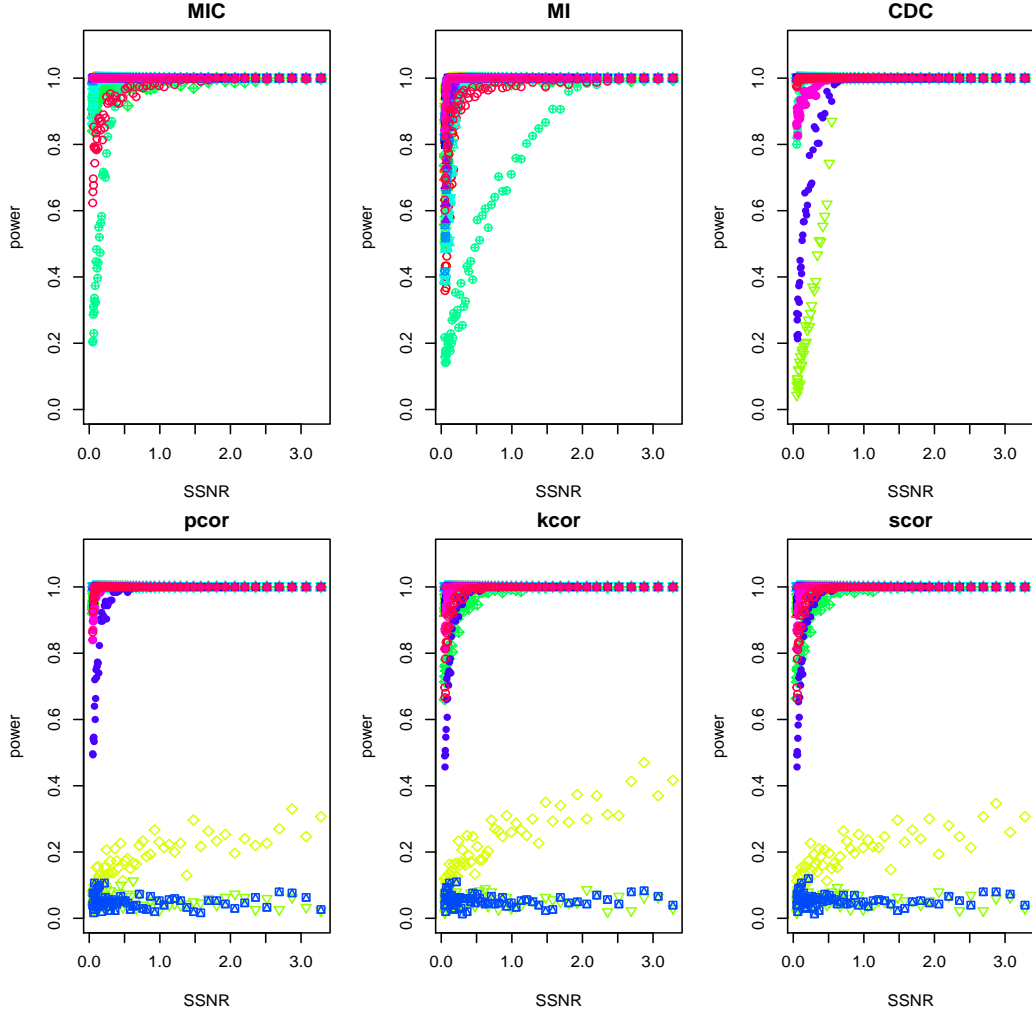


Figure 8: (color online) Weak-Equitable of MIC, MI, CDC, pcor, kcor and scor with mean SSNR range from 0.0496 to 3.282. (Legends are the same as that in Figure 1.) The simulation is based on 21 types of relationship (Definitions are given in appendix), and SSNR is controlled by the heuristic algorithm 1 and eq. 7. In this simulation we take the sample x from uniform distribution with length 5000, and noise is sampled from standard normal distribution. The power was estimated from 300 times simulation at significant level $\alpha = 0.05$. We can see from the results that CDC and MIC are much more equitable than others in that range of SSNR. For MI and MIC, the power-equitable of MI is similar than that of MIC.

- [5] Ruth Heller, Yair Heller, Malka Gorfine. A consistent multivariate test of association based on ranks of distances. arXiv:1201.3522v1.
- [6] C.E.Shannon. A mathematical theory of communication. Mobile computing and communication review, Volume5, Number1
- [7] R.Steuer, J.Kurths, C.O.Daub, J.Weise and J. Selbig,(2002) The mutual information: Detecting and evaluating dependencies between variables. BIOINFORMATICS. Vol.18 Suppl.2.S231-S240
- [8] E.H.Linfoot. An information measure of correlation. Information and Control,1(1957),p85-89
- [9] Young-II Moon, Balaji Rajagopalan, and Upmanu Lall. Estimation of mutual information using kernel density estimators. PHYSICAL REVIEW E. VOLUME 52. NUMBER 3. 1995.
- [10] David N.Reshef, et al. Detecting Novel Associations in Large Data Sets. Science.334,1518.
- [11] Noah Simon and Robert Tibshirani. COMMENT ON "Detecting Novel Associations in Large Data Sets"
- [12] Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning. Springer.2001
- [13] Delicado P, Smrekar M. Measuring non-linear dependence for two random variables distributed along a curve[J]. Statistics and Computing, 2009, 19(3): 255-269.
- [14] Yin X. Canonical correlation analysis based on information theory[J]. Journal of Multivariate Analysis, 2004, 91(2): 161-176.
- [15] Lopez-Paz, David, Philipp Hennig, and Bernhard Schölkopf. "The randomized dependence coefficient." Advances in Neural Information Processing Systems. 2013.
- [16] Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, Alexander Smola. A Kernel Two-Sample Test. Journal of Machine Learning Research. 13(2012):723-773.
- [17] Pczos B, Ghahramani Z, Schneider J. Copula-based kernel dependency measures[J]. arXiv preprint arXiv:1206.4682, 2012.
- [18] Renyi A. On measures of entropy and information[C]//Fourth Berkeley Symposium on Mathematical Statistics and Probability. 1961: 547-561.
- [19] Tsallis,C. Possible generalization of boltzmann-gibbs statistics. J. Statis. Phys. 52(1-2):479-487, 1988
- [20] Fernandes A D, Gloor G B. Mutual information is critically dependent on prior assumptions: would the correct estimate of mutual information please identify itself?[J]. Bioinformatics, 2010, 26(9): 1135-1139.
- [21] Khan, Shiraj; Bandyopadhyay, Sharba; Ganguly, Auroop R.; Saigal, Sunil; Erickson, David J. III; Protopopescu, Vladimir; and Ostrouchov, George, "Relative performance of mutual information estimation methods for quantifying the dependence among short and noisy data" (2007). Civil and Environmental Engineering Faculty Publications.
- [22] Alexander Kraskov, Harald Stogbauer, and Peter Grassberger. Estimating Mutual Information.
- [23] Tan Q, Jiang H, Ding Y. Model selection method based on maximal information coefficient of residuals[J]. Acta Mathematica Scientia, 2014, 34(2): 579-592.
- [24] Nelsen R B. An introduction to copulas[M]. Springer, 1999.
- [25] Bach, Francis R. Kernel independent component analysis. JMLR, 3:1-18,2002
- [26] Gretton, A., Herbrich R., and Smola, A. The kernel mutual information. In Proc. ICASSP, 2003.
- [27] Gretton,A., Bousquet, O., Smola, A. and Scholkopf, B. Measuring statistical dependence with Hilbert-Schmidt norms. In ALT, pp. 63-77, 2005.

- [28] B. Schweizer and E.F. Wolff. On nonparametric measures of dependence for random variables. The Annals of statistics, 1981, Vol.9, No.4, 879-885.
- [29] Amir Dembo, Abram Kagan, Lawrence A. Shepp. Remarks on the Maximum Correlation Coefficient. Bernoulli, Vol.7, No.2 (Apr., 2001), pp. 343-350
- [30] Nickos Papadatos, Tatiana Xiafara. A simple method for obtaining the maximal correlation coefficient and related characterizations. Journal of Multivariate Analysis 118(2013) 102-114.
- [31] Wlodzimierz Bryc, Amir Dembo, Abram Kagan. On the maximum correlation coefficient. Technical Report No. 2002-25 August 2002.
- [32] Masashi Sugiyama, Makoto Yamada. On kernel parameter selection in Hilbert-Schmidt Independent Criterion. IEICE Transactions on Information and Systems, Vol.E95-D, no.10, pp.2564-2567, 2012
- [33] Arthur Gretton, Ralf Herbrich, Alexander Smola et al. Kernel Methods for Measuring Independence. Journal of Machine Learning Research 6(2005) 2075-2129.
- [34] Dino Sejdinovic, Bharath Sriperumbudur, Arthur Gretton, Kenji Fukumizu. Equivalence of distance-based and RKHS-based statistics in hypothesis testing. arXiv: arXiv:1207.6067
- [35] Charles A. Micchelli, Yuesheng Xu, Haizhang Zhang. Universal kernels. Journal of Machine Learning Research 7(2006) 2651-2667.
- [36] Reshef DN, Reshef Y, Mitzenmacher M, Sabeti P (2013) Equitability analysis of the maximal information coefficient with comparisons arXiv:1301.6314v1
- [37] Justin B. Kinney and Gurinder S. Atwal Equitability, mutual information, and the maximal information coefficient PNAS 2014 111 (9) 3354-3359;
- [38] Hangjin Jiang, Yiming Ding. Dependence Measure: A comparative Study.
- [39] Reshef, David N., et al. "Cleaning up the record on the maximal information coefficient and equitability." Proceedings of the National Academy of Sciences 111.33 (2014): E3362-E3363.

5 Appendix

5.1 Definition of functions

1. Line: $y = x$
2. Linear+Periodic, Low Freq: $y = 0.2 \sin(4(2x - 1)) + \frac{11}{10}(2x - 1)$
3. Linear+Periodic, Medium Freq: $y = \sin(10\pi x) + x$
4. Linear+Periodic, High Freq: $y = 0.1 \sin(10.6(2x - 1)) + \frac{11}{10}(2x - 1)$
5. Linear+Periodic, High Freq: $y = 0.2 \sin(10.6(2x - 1)) + \frac{11}{10}(2x - 1)$
6. Non-Fourier Freq [Low] Cosine: $y = \cos(7\pi x)$
7. Cosine, High Freq: $y = \cos(14\pi x)$
8. Cubic: $y = 4x^3 + x^2 - 4x$
9. Cubic, Y-stretched: $y = 41(4x^3 + x^2 - 4x)$
10. L-shaped: $y = x/99I(x \leq \frac{99}{100}) + I(x > \frac{99}{100})$
11. Exponential $[2^x]$: $y = 2^x$
12. Exponential $[10^x]$: $y = 10^x$

13. Parabola: $y = 4x^2$
14. Non-Fourier Freq [Low] Sine: $y = \sin(9\pi x)$
15. Sine, Low Freq: $y = \sin(8\pi x)$
16. Sine, High Freq: $y = \sin(16\pi x)$
17. Sigmoid: $y = [50(x - 0.5) + 0.5]I(\frac{1}{20} \leq x \leq \frac{51}{100}) + I(x > \frac{51}{100})$
18. Varying Freq [Medium] Cosine: $y = \sin(5\pi x(1 + x))$
19. Varying Freq [Medium] Sine: $y = \sin(6\pi x(1 + x))$
20. Spike: $y = 20I(x < \frac{1}{20}) + [-18x + \frac{19}{10}]I(\frac{1}{20} \leq x < \frac{1}{10}) + [-\frac{x}{9} + \frac{1}{9}]I(x \geq \frac{1}{10})$
21. Lopsided L-shaped: $y = 200xI(x < \frac{1}{200}) + [-198x + \frac{199}{100}]I(\frac{1}{200} \leq x < \frac{1}{100}) + [-\frac{x}{99} + \frac{1}{99}]I(x \geq \frac{1}{100})$

where $I(\cdot)$ is the indicator function.

5.2 Data for simulation

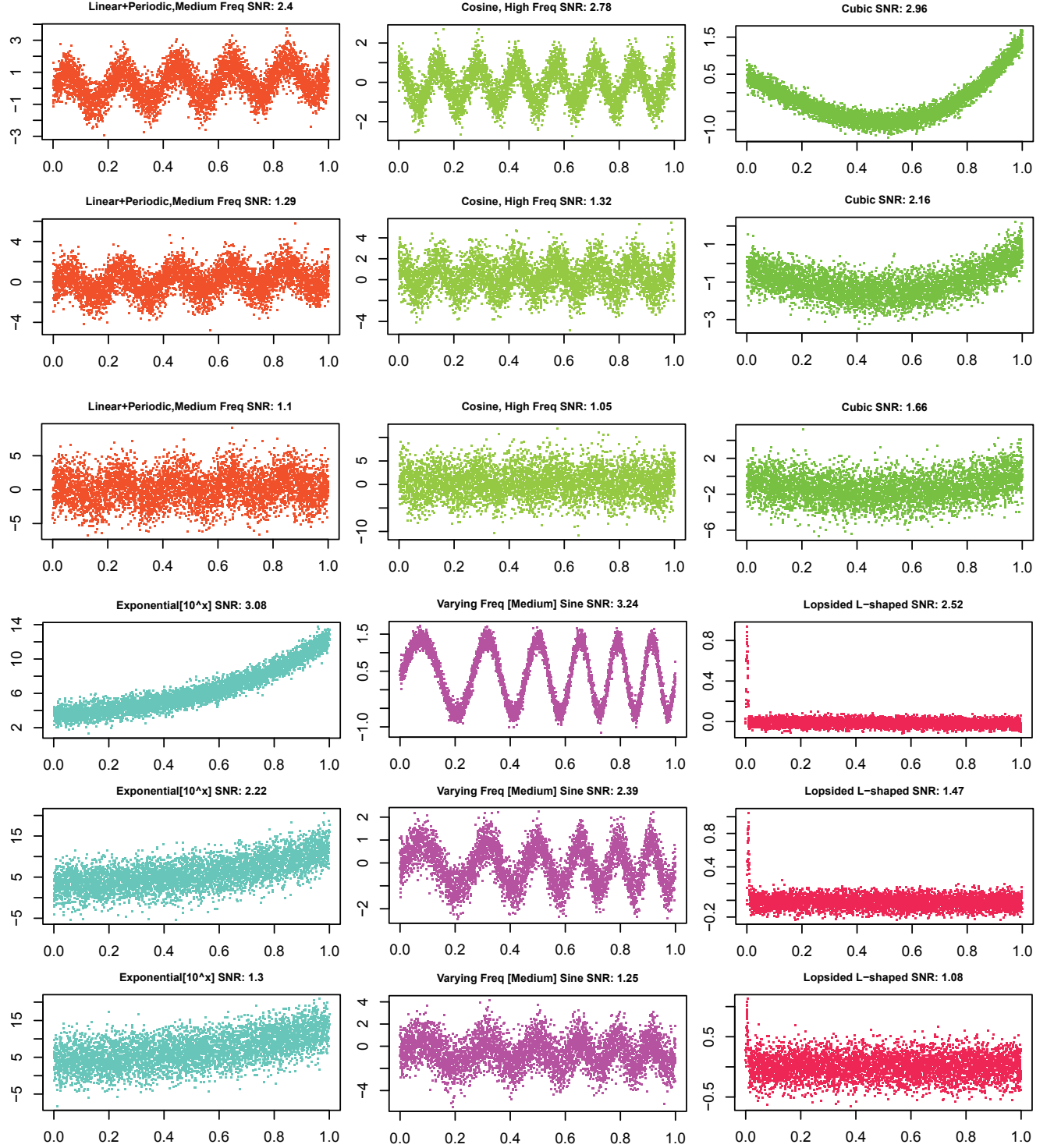


Figure 9: 6 types of relationships with different SNR are shown here. Definitions of relationships are given in appendix 5.1), and SSNR is controlled by the heuristic algorithm 1 and eq. 7. All the simulations about SSNR were taken by such data.